

Review

of the thesis of Dr. Velislava Stoykova
“Machine Translation Approaches to the Bulgarian Language”,
presented for awarding the scientific degree “doctor of science”
in Professional area 2.1 Philology

by Prof. Dr. Maria Nisheva-Pavlova – Faculty of Mathematics and Informatics,
Sofia University St. Kliment Ohridski

Pursuant to Order 3-PK-268/13.07.2022 of the Rector of NBU I am nominated as a member of the scientific jury for the defense of the submitted thesis in professional area 2.1 Philology, entitled “Machine Translation Approaches to the Bulgarian Language”. At the first meeting of the scientific jury, I was designated as a reviewer for this procedure.

1. General characteristics of the dissertation and the presented materials

The dissertation is written in English. It contains 110 pages of text and is divided into two parts, including a total of five chapters, a list of references, two appendices, a glossary and an index.

In addition to the dissertation on the procedure, the following are also presented:

- an abstract in Bulgarian, containing 43 pages of text;
- a European-style CV.

Copies of the publications related to the dissertation and a list of noted citations of these publications were not submitted, but such are not required by the Ordinance on the Development of the Academic Staff of the NBU.

2. Applicant data

Dr. Velislava Stoykova has Master’s degree in Bulgarian language and literature, acquired at the Faculty of Slavic Philology of Sofia University St. Kliment Ohridski. In 2003, she defended a PhD thesis on the topic “Formal morphological analysis of the

definiteness/indefiniteness category in Bulgarian, modeled in DATR – theory”. Since 2011, he has held the academic position of chief assistant professor at the Institute for Bulgarian Language at the Bulgarian Academy of Sciences.

I know Dr. Stoykova vaguely from some of her participations in international scientific conferences, where we had opportunities to make professional contacts. I have very good impressions of her drive to learn new technologies and carry out in-depth interdisciplinary research.

3. Relevance of the research area and significance of the research problem

The dissertation is devoted to research in the field of language technologies. It presents the author’s results of the study of various formal linguistic theories and their applicability for the purposes of automatic analysis of the Bulgarian language and machine translation into Bulgarian. Results of the application of numerous statistical techniques for machine translation into Bulgarian language are discussed as well.

The area of research in the doctoral thesis is complex and modern, and its relevance is determined by the expanding set of real and potential applications of machine translation, in particular machine translation into the Bulgarian language.

The goal of the dissertation and the tasks for achieving this goal, which the author indicates implicitly both in the dissertation and in the abstract, are determined on the basis of a brief analysis of the achievements and open questions in the chosen field. The goal is meaningful and the tasks are aligned with the goal.

4. Analysis of the content, results and contributions of the doctoral thesis

The dissertation consists of 110 pages of English text, including five chapters, list of references and two appendices.

The introductory first chapter presents the field of research and basic concepts in this field. The main machine translation approaches that are of practical importance – rule-based machine translation and statistical machine translation – are discussed.

The second chapter is devoted to study of the applicability of the lexical information representation language DATR for machine translation purposes. The structure of the representation of grammatical knowledge by means of DATR and the existing software tools supporting machine translation capabilities between languages whose grammatical rules are described in DATR terms are succinctly described. The results of the author's in-depth analysis of the possibilities of using DATR to represent the word changes of nouns in Bulgarian and Russian for the purposes of machine translation are discussed. Special attention is paid to the use of DATR as a transformational grammar to compare the representation of possessive and reflexive-possessive pronouns in the Bulgarian language.

Chapter three addresses a series of issues related to the use of the Universal Network Language UNL for machine translation purposes. The structure of UNL and the purpose of the basic language constructs are described. The possibilities of using UNL to represent the word changes of nouns in the Bulgarian language have been analyzed and a comparison with the corresponding possibilities provided by DATR has been made. The results of the author on the presentation of the possessive and reflexive-possessive pronouns in Bulgarian, English and Russian by means of UNL are thoroughly discussed. The discussion is supported by relevant examples. The most significant contribution of this chapter is the presentation of results related to the use of UNL for machine translation between Bulgarian and Slovak. Opportunities for using UNL to measure and improve the quality of machine translation are also discussed.

The fourth chapter is dedicated to a study of the machine translation capabilities using the Sketch Engine (SE) search engine. The application of SE for machine translation of terms from the field of ecology between Bulgarian and Slovak has been analyzed. Attention has been paid to using SE for machine translation of time expressions between Bulgarian and Slovak languages. The presentation of this chapter focuses on the description of the theoretical analysis and practical experiments carried out, aimed at using SE for machine translation between Bulgarian and Slovak languages. Possibilities of using SE for machine translation of mathematical terms between Bulgarian and Serbian have also been demonstrated.

The concluding fifth chapter contains summaries of the results presented in the previous three chapters and the corresponding conclusions of the author, as well as summarizing conclusions of the author's research on the topic of the dissertation.

The list of references includes 105 titles. The manner of citing them in the text of the dissertation testifies to a very good knowledge of the state of research and applied developments in the chosen field.

The appendices contain a number of figures illustrating in a suitable way the presentation of some of the sections of Chapter 2 and Chapter 3.

The main contributions of the doctoral thesis of Velislava Stoykova may be summarized as follows:

- A comparative analysis of existing software tools for machine translation into Bulgarian, based on the use of formal linguistic theories, was performed.
- An original technique for measuring the quality of the result of machine translation into Bulgarian is proposed and possibilities for its improvement by improving the formal presentation of language rules are formulated.
- An electronic corpus of specialized mathematical texts was created and used for statistical machine translation. Some promising results of such a translation between Bulgarian and Serbian have been obtained.
- Results have been obtained in establishing the limitations of the use of machine translation in Bulgarian as a complement to and supporting the translation performed by humans.

The dissertation makes an excellent impression with the scope and depth of its presentation. The field of research is modern and complex, and achieving significant results in it requires serious interdisciplinary knowledge and skills, constant and intensive work. The achieved results are original and significant and fully correspond to the defined objective.

5. Publications on the doctoral thesis. Reflection on the works of other authors

The results obtained in the dissertation are presented in a total of 12 papers, 11 of which are published as follows:

- five – in scientific editions, referenced and indexed in WoS and/or Scopus;
- four – in proceedings of international scientific conferences;
- two – in a series, referenced and indexed in Google Scholar.

A paper is about to be published in a Springer edition that is refereed and indexed in Scopus.

Eight papers from this list are single-authored and four are co-authored. I suppose that all co-authors have contributed equally to each of the collective publications.

No data is presented on noticed citations of Velislava Stoikova's publications based on the results of her dissertation in papers by other authors, but a total of four such citations were found in a Scopus search, and more than 15 such citations – in Google Scholar (excluding self-citations).

In this way, the requirements of Art. 1a of the Regulations for the Implementation of the Law on the Development of the Academic Staff in the Republic of Bulgaria for the acquisition of a scientific degree “doctor of sciences” in professional area 2.1 Philology have been fulfilled and significantly exceeded.

6. Critical Remarks and Recommendations

In general, the dissertation is designed carefully and correctly. The presentation is competent and sufficiently thorough. As more significant shortcomings in terms of content, I would point out the lack of a clear formulation of the tasks of the presented research and the lack of sufficient argumentation of the choice of the used linguistic and software tools to achieve the goal of the research. It would also be appropriate to include a more extensive and complete overview of the state of scientific research and practical results in the field of machine translation, accompanied by the formulation and analysis of the most important open questions in this field, some of which are explored in the dissertation.

However, I suppose that most of the mentioned shortcomings are primarily due to the fact that the dissertation is in the form of a book submitted to a world-renowned publisher, and the author had to comply with the corresponding requirements.

7. Abstract

The abstract meets all the requirements for its preparation and fully and accurately presents the topic, purpose, content, achieved results and contributions of the dissertation.

8. Summary

Summing up, I consider that the dissertation of Velislava Stoykova fully satisfies and significantly exceeds the requirements of the national regulations and the specific conditions and requirements of the New Bulgarian University. Its author has achieved significant research results that make an original contribution to the chosen field of study. My assessment of the dissertation, the abstract and the scientific contributions of their author, Velislava Stoykova, is **positive**.

Therefore, **I advise the honorable scientific jury to award to Velislava Stoykova the scientific degree “doctor of science” in professional area 2.1 Philology.**

Sofia, September 8, 2022

Prof. Dr. Maria Nisheva-Pavlova